DOCUMENT RESUME

ED 470 296 TM 034 518

AUTHOR	Easton, John Q.
TITLE	Testing and Assessment in Illinois School Districts.
INSTITUTION	Consortium on Chicago School Research, IL.
PUB DATE	2000-11-00
NOTE	10p.
PUB TYPE	Reports - Research (143)
EDRS PRICE	EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS	Academic Achievement; Educational Assessment; Elementary Secondary Education; *School Districts; *Standardized Tests; Surveys; *Test Use; *Testing Programs
IDENTIFIERS	*Chicago Public Schools IL; *Illinois

ABSTRACT

A survey of school districts across the state of Illinois in 1999 focused on three dimensions of local assessment practices: (1) why districts give tests; (2) the tests they give; and (3) what they do with test results. Telephone interviews were conducted with district assessment coordinators or superintendents from 75 school districts, including the state's largest school districts. Nearly 90% of districts mentioned student assessment purposes as one of the major purposes of their district-wide testing program. More than 70% of districts stated that measuring student performance is the major purpose of their testing program. Slightly fewer than one quarter of districts mentioned that a major purpose of the testing program is to compare students, schools, and districts to national norms as an external check. Almost 30% of districts identified student placement as a major purpose of the assessment program. Curriculum and program evaluation was mentioned by 56% of districts. Reporting of results and planning and goal setting were also mentioned by many districts. About 90% of districts administer standardized achievement tests to their students, and the next most prevalent type of assessment is tests of aptitude or intelligence. Relatively few districts administered locally developed assessments. Districts identified strengths and weaknesses in their own testing programs, with lack of alignment with curriculum the most commonly cited weakness. Districts tended to place a lot of faith in the standardized assessments they used, but were less enthusiastic about the Illinois Standards Achievement Testing (ISAT) program, believing the ISAT to be of lower quality than the assessments they had chosen. (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 470 296

Testing and Assessment in Illinois School Districts

John Q. Easton

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

J.Q. Easton

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM034518

Consortium on Chicago School Research

BEST COPY AVAILABLE

Testing and Assessment in Illinois School Districts

November 2000 John Q. Easton

Study Introduction

The Consortium on Chicago School Research (CCSR) and the Illinois Business Roundtable (IBRT) have historic interests in uses of student assessment information. The two organizations came together in fall 1999 to plan and then conduct a survey of representative school districts across the state. The survey focused on three dimensions of local assessment practices within the districts: why districts give tests, what tests they give, and what they do with their test results? Finally, the survey solicited responses from districts about the Illinois Standards Achievement Tests (ISAT) program, how it could be improved, and how it could better meet districts needs.

Procedures

The Consortium drew a random sample of 60 districts across the state, drawing proportionately from metropolitan and downstate districts and assuring representation of small, medium, and large-sized districts. In addition, CCSR identified the 20 largest school districts in Illinois. With the assistance of Research Partnerships of Wheaton, Illinois, telephone interviews were conducted with district assessment coordinators or superintendents. Principal researchers from CCSR and IBRT interviewed representatives from the largest districts and staff from Research Partnerships interviewed the remaining districts. The interviews lasted between 30 and 45 minutes. Seventy-five districts completed the assessment survey, providing a fair representation of district testing practices statewide.

Results

Purposes of District Assessment Programs

The majority of the survey questions focused on why district-wide tests are administered and how the results are used. In the very first question, we asked the testing administrator or other appropriate staff member to tell us the major purposes of the district testing program. Typically, each district mentioned two to four different major purposes. Not at all surprising, the most common set of responses related to student assessment purposes. **Nearly 90 percent of districts mentioned student assessment purposes as one of the major purposes of their district-wide testing program.** See Figure 1 for a display of these results.

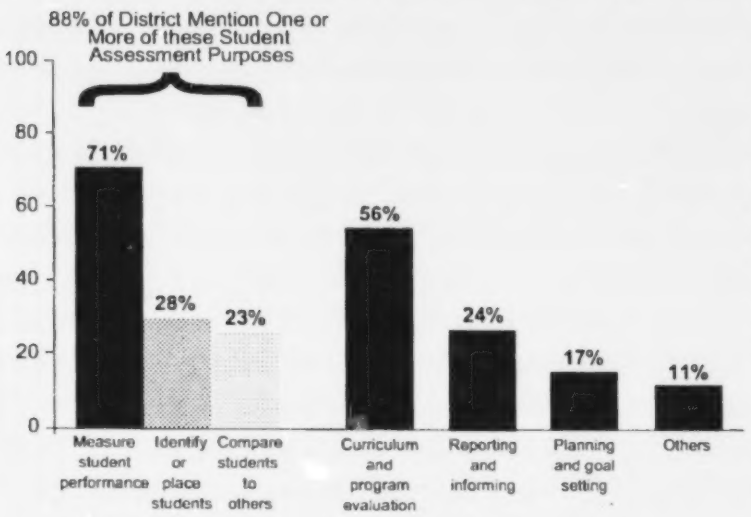
This overall purpose can be best understood by looking at various different subcategories within this umbrella category. **Over 70 percent of the districts stated**

The Consortium on Chicago School Research 1313 East 60th Street Chicago, Illinois 60637
773-702-3364 • 773-702-2010 - fax • **www.consortium-chicago.org**

BEST COPY AVAILABLE

Figure 1

Major Purposes of District Wide Testing Programs



that measuring student performance is the major purpose of their testing program. The general idea expressed by these districts is that test scores provide information about the overall level of student performance, much like a thermometer or speedometer. The test scores identify overall student strengths and weaknesses, they provide feedback relative to district and/or state goals, they provide external evidence of student improvement and growth, and they give "a general reading of how students are doing in the instructional program." In a closely related category, **slightly fewer than one-quarter of districts specifically mentioned that a major purpose of their testing program is to compare their students, schools and district to national norms as an external check.**

Still within the category of student assessment is a separate grouping of responses that refer to using assessment results to **identify or place students** in special programs or to refer them for particular instruction. These programs run the range from special education to gifted programs. Identifying low achieving students and students with other difficulties in order to provide them with needed services, placing students in appropriate classes, and pre-screening for learning disabilities are included here. **Almost 30 percent of districts described using test scores to identify or place students as a major purpose of their assessment program.**

The second most common purpose is curriculum and program evaluation, which was mentioned by 56 percent of districts. Districts use the results of their assessment programs to assess and obtain feedback on their curriculum needs, to target curriculum areas for improvement, and to adjust and fine-tune curriculum sequence and scope. Districts also describe using assessment results to provide feedback on instruction and to use this information for instructional improvements. This process occurs by providing assessment results in terms of areas of strengths and weaknesses. Also, within this category are uses related to program evaluation. Districts report using their test results to review specific programs and to monitor their effectiveness.

The next major purpose of district assessment, cited by 24 percent of districts, is for **reporting of results** to parents, the public, and the school board. Districts rely on test results to inform parents how well their students are doing academically. They also use the test information to inform the broader community about the quality of the district's education program. Reporting test results to the local board of education was also mentioned here. In all cases, these uses are related to making the district publicly accountable to a variety of important stakeholders.

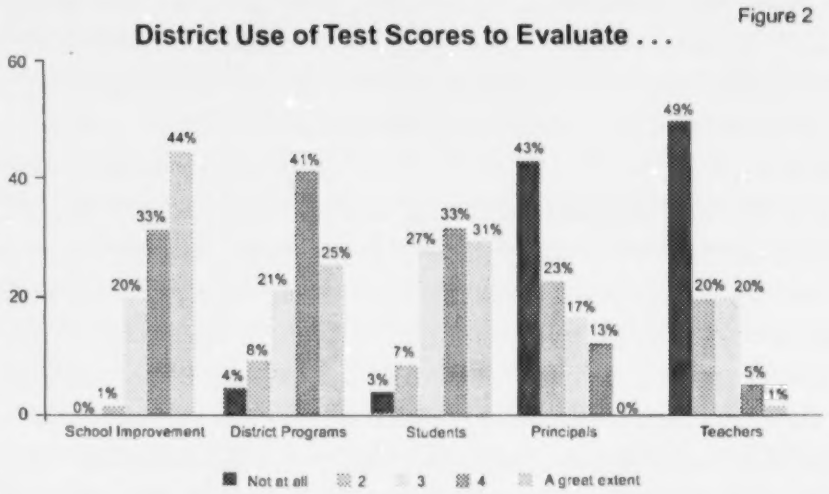
The fourth major testing purpose can be described as **planning and goal setting.** Seventeen percent of responding districts mentioned these activities as an important purpose of their assessment program. Several districts describe using assessment results for school improvement planning and use in continuous quality improvement. Other related uses include setting annual goals and then reviewing test scores in that context.

The purposes described above were noted in response to an open-ended question. They can be compared to responses to a forced-choice question in which the district rated the use of test scores in five areas using a five-point scale where 1 represents "not at all" and 5 equals "a great extent." The districts were asked to rate the extent to which test score results are used to evaluate district programs, school improvement, principals, teachers and students (see Figure 2).

BEST COPY AVAILABLE

In terms of actual usage of test score results, district rate "evaluating school improvement" higher than for any of the other areas. About 77 percent of districts chose "to a great extent" or the next highest category to describe using test score results to evaluate school improvement. About two-thirds of districts use the two highest categories to describe their use of test score to evaluate district programs. Using test scores to evaluate students received similar ratings. It is notable that whereas districts offer student assessment related purposes as the predominant reasons they have their testing program, in actual ratings of usage they report more use of assessment results for evaluating school improvement and district programs than students.

Figure 2



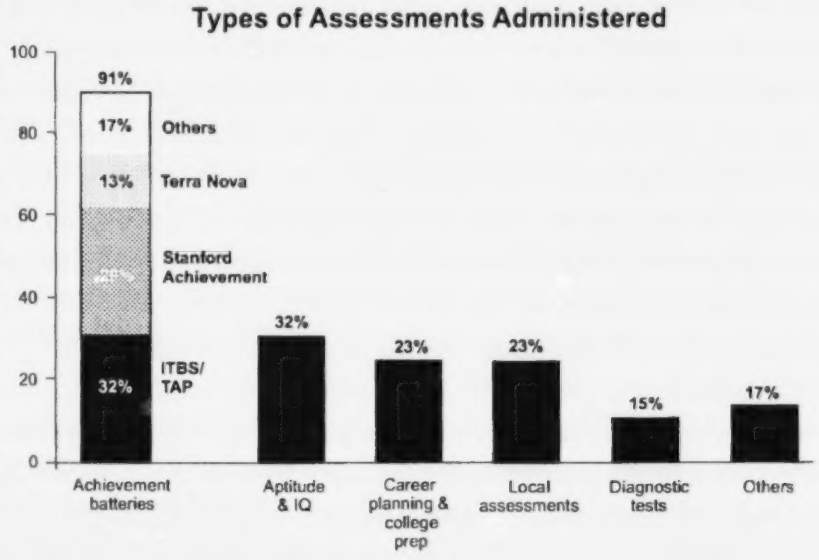
In contrast to evaluating schools, programs and students, few districts use test scores extensively to evaluate either principals or teachers. In both cases, the most frequent response is "not at all."

Types of Assessments Administered

Given the many different purposes that districts have for administering assessments, it is no surprise that districts use a great variety of types of assessments. Most of these are commercially produced "off the shelf" products, created for general testing purposes. By far the largest group of these consists of **achievement tests**, including the Iowa Tests of Basic Skills (with the Tests of Achievement Proficiency for high school students), the Stanford Achievement Tests, the Terra Nova, plus others. **About 90 percent of districts administer standardized achievement tests to their students** (see Figure 3).

The next most prevalent type of assessment is tests of student **aptitude or intelligence**. These tests are most often used for placing students in gifted or remedial programs. The two most commonly reported tests include the Otis-Lennon School Ability Tests (OLSAT), which measures cognitive abilities and can be used to compare student ability to achievement. The OLSAT is designed for use in conjunction with the Stanford Achievement Tests. The second ability test used in many districts is the Cognitive Abilities Test (CogAT). This test is meant to assess students' abilities in reasoning. Because it is published by the

Figure 3



BEST COPY AVAILABLE

same company that sells the Iowa, the two can be used together to compare ability and achievement. Approximately one-third of districts administer these aptitude or IQ tests to their students, though usually only to selected grades.

About 23 percent of districts administer **career planning and college preparatory** instruments to students. These are typically given to high school students or eighth grade students. The most common of these tests is the ACT PLAN, developed and distributed by the American College Testing Program. It consists of both a set of achievement tests and non-academic sections including an interest inventory, and educational and occupational plans. Students, parents, and counselors use the results for planning post-secondary endeavors and for helping with course selection in the final two years of high school. The ACT EXPLORE is a similar test for eighth grade students who may use the results in planning their high school programs.

About an equal number of school districts have created their own **local assessments,** aligned with the district curriculum. These are often called CRTs—for criterion-referenced tests. Districts use these tests for more immediate feedback about student progress through the local curriculum. These assessments are often described as "curriculum embedded" and provide information specific to the district instructional program.

Finally, 15 percent of districts administer diagnostic tests, most frequently to students in primary grades. The two most used of these tests are the Gates MacGinities Reading Test and the Developmental Reading Assessment. These tests are administered to provide detailed, in-depth information about students strengths and weaknesses, with instructional implications for improvement.

More About Achievement Tests: Grades Tested, Time on Testing, and Cost

Nearly every district that responded to this survey administers a standardized achievement battery in some grades. In elementary grades these tests are most typically administered in reading, math, science and social studies, beginning in grade two or three (though more than half of districts also test first graders), through eighth grade. Grades three through eight are the most tested grades, with between 90 and 100 percent of districts giving assessments in these grades. In high schools, on the other hand, about one-half of districts administer achievement tests to students in grades nine and eleven, with somewhat more testing tenth graders. Twelfth grade achievement testing is rare. Most districts test in either fall or spring (these two times are equally popular) though about 20 percent test in winter. A few districts test both fall and spring in order to measure growth within the school year.

The annual testing time required for achievement batteries ranges from a low of two hours in districts that test only math and reading to a high of six to eight hours. Districts with the greatest amount of testing time assess more subjects, including writing.

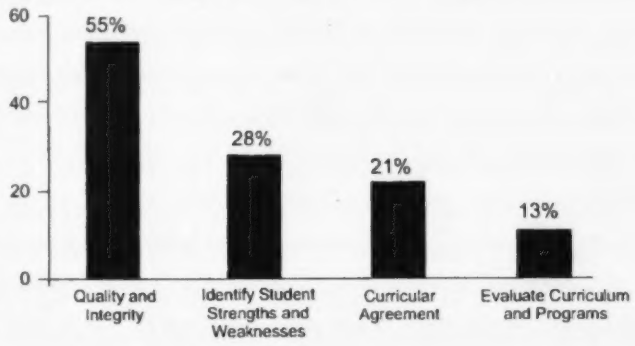
Districts had some difficulty in estimating the total cost of their testing programs. The average estimate, however, was in the range of $11 to $15 per student

Strengths of District Testing Programs

The districts noted numerous strengths with their testing program (see Figure 4). More than half of them described strengths in terms of the **Quality and Integrity of the Testing Program**. Many attributes contribute to the overall quality. The most frequent comments emphasized the consistency, objectivity, fairness, accuracy, and credibility of the testing program. Almost one-quarter of districts used one of these specific terms in describing their strengths. They also said that the tests had "very solid reputations," they

Figure 4

Strengths of District Programs



were "widely used," and importantly, "we have faith in it."

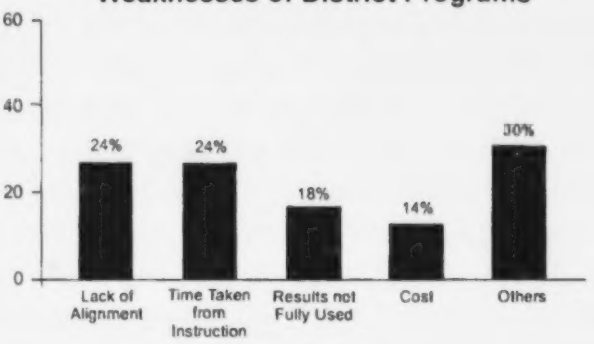
Another important aspect of the perceived high quality of the testing programs is the ability to track trends from one year to the next. Nine districts stated their ability to compile historical data to examine trends contributed to the strength of the testing system. An equal number described the importance of national norms, because they show "where students are on a national level" and that the "national comparison gives us a broader perspective." Finally, contributing to the quality of the assessment programs, were comments about the attractiveness of the testing materials, and the support for the program from teachers and parents.

The second largest category of strengths concerns how assessment programs help to **Identify Student Strengths and Weaknesses**. Districts made comments like, "the tests provide an accurate measure of how students are doing," we "can tell if students need extra help" and they identify "student strengths and weaknesses to allow us to better meet their needs." Twenty-eight percent of districts used similar language to describe strengths of their testing programs.

The next most prevalent responses is related to **Curricular Alignment**. Twenty-one percent of districts made comments about alignment between their test and their curriculum, instruction or learning standards. Of these, a small number use a standardized test that is specifically designed to measure the Illinois Learning Standards. These districts saw this alignment to state learning standards as a strength of their testing program. More generally, respondents commented that "the tests are as closely aligned to the curriculum as possible, it's a pretty good match to our instructional program," and "it covers areas that are important." These districts acknowledged that while their testing program may not be strictly aligned to state learning standards they have confidence that they are measuring the same important expectations.

Figure 5

Weaknesses of District Programs



The final major category of strengths, noted by 13 percent of districts, is their ability to **Evaluate Curriculum and Programs**. Testing results help districts to "evaluate the strengths and weaknesses of the curriculum," they "strengthen curriculum decisions," and they "identify curriculum areas that need addressing." Districts noted several other strengths, including the ability to communicate with parents and teachers, and that their assessment programs provided them with a variety of different measures of students' progress.

Weaknesses of District Testing Programs

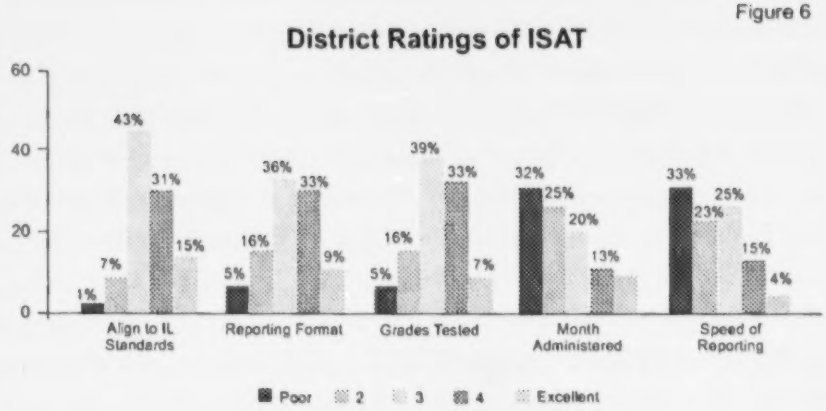
The most frequently stated weakness of district assessment programs, noted by 24 percent of respondents, is **Lack of Alignment** with learning standards, curriculum, and instruction (see Figure 5). Comments reported by districts include: "it doesn't always match what we are teaching and our classroom practices,"

We are able to isolate the relatively few districts that administer locally developed tests and analyze their strengths and weaknesses separately from other districts. The strengths of these include being curriculum-based and aligned so that they meet the needs of their students. The weaknesses include not being professionally developed and the difficulties of charting trends and disaggregating results by groups of students.

BEST COPY AVAILABLE

"there is not a perfect alignment to curriculum" and "it doesn't always measure what our curriculum teaches." Several districts used virtually identical phrases: "not tied directly to Illinois Learning Standards." One district said that the results are "based on someone else's norm group—what exactly is that?"

Figure 6



An equal number of districts—24 percent—pointed to the amount of **Time Taken from Instruction** as a problem in their testing programs. Administering tests takes teacher time and student time, with the net effect that less time is available for instruction. In a related vein, 14 percent of respondents mention high **Cost** as a weakness.

A final category, noted by 18 percent of districts, is **Results Not Used to Full Potential**. These responses focused on problems in interpreting results, the need for additional training in using test results, the time needed to analyze results, and possible misinterpretation by non-educators.

Finally, there were a number of other responses that did not fit as neatly into categories. Six districts said that students either do not take the tests as seriously as they should or that there is too much stress associated with testing. Several districts noted the need for more diverse assessments and fewer multiple-choice assessments. Other districts noted shortcomings that were within their own ability to remedy (e.g., time of year tests administered, tests too easy for student population).

Rating the ISAT

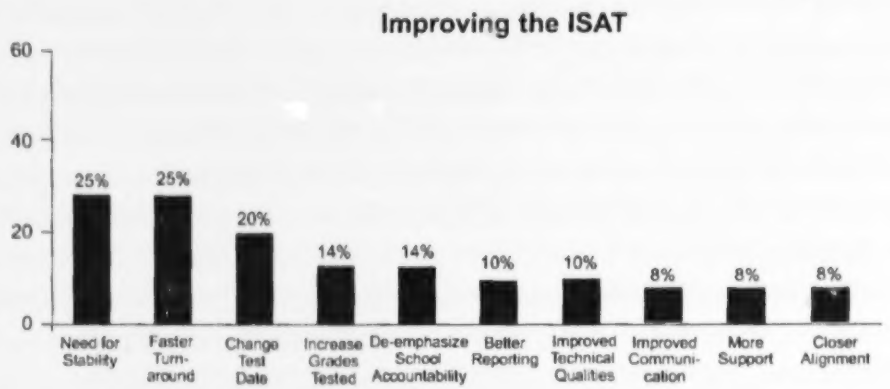
Though most of the questions on the survey dealt with district or local assessment programs, the final questions focused on the Illinois Standards Achievement Testing (ISAT) program. The first set of these asked districts to rate the ISAT program using a five-point scale ranging from poor (given a value of 1) to excellent (given a value of 5). Districts rated the ISAT program on how well it is aligned to the Illinois Learning Standards, month administered, speed of reporting results, reporting format, and grades tested (see Figure 6). Among these five areas, districts gave the highest rating to alignment to learning standards, though in absolute terms even this item does not receive a very high rating. Fewer than half of the districts use category 4 or 5 in rating the alignment of ISAT to state learning standards. The ISAT reporting format and the grades tested also receive relatively high ratings, with 41 and 40 percent respectively of districts using the two highest categories. The final two items—the month that the ISAT is administered (which was February 2000) and speed of reporting—both receive quite low ratings. Only 21 and 19 percent respectively use the two high categories. In both cases, the most frequent rating for these two items is "poor," the lowest possible rating.

Improving the ISAT

A final, open-ended question on the survey asked districts to describe how the ISAT program could be improved (see Figure 7). The most frequent response made by about one-quarter of districts was the **Need for Stability and Consistency**. Of these comments, nearly all used the specific words "consistency" and "stability." Districts reported that changes in test format and grades tested are disruptive to districts and make the possibility of tracking trends over time difficult, if not impossible. There is a very vocal desire among the responding districts for the State Board of Education to "make a plan and stick to it."

BEST COPY AVAILABLE

Figure 7



An equal number of districts (25 percent) urged **Faster Turn-around.** It takes "way too long" to get test results back. Twenty percent of districts asked to **Change the Test Date** to later in the school year. However, a few districts requested early fall testing.

A number of districts (14 percent) advocated to **Increase the Number of Grades Tested**. One rationale for the increased testing is that if adjacent grades are tested, then test score gains can be calculated. Gain scores provide the foundation for "value-added" measures of school improvement. Another rationale is that with more grades tested, districts might be able to reduce their own testing.

Fourteen percent of districts suggested that ISBE **De-emphasize Accountability and Emphasize School Improvement** aspects of the assessment program. They expressed some frustrations with the use of test results to compare schools to each other, at the expense of providing useful information for improvement activities. Interestingly in this context, several districts (about eight percent) think that ISAT would be improved by making the testing system **High Stakes for Students.** They believe the test should be made to pressure students to perform and achieve better.

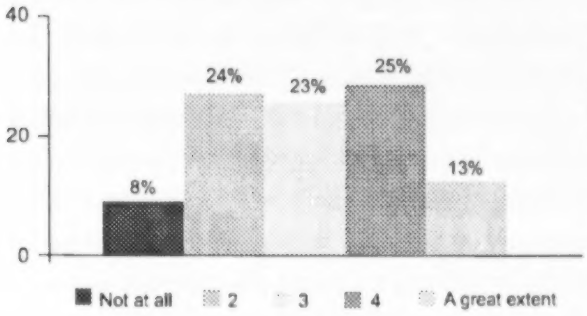
Between eight and 10 percent of districts suggested that ISAT would be improved by making the following changes. **Better Score Reporting** includes that the test results take into account the background of the students, that results be reported via computer, that additional item analyses are included, and that both national and international comparisons be made. For **Improved Communication**, districts requested better coordination with teachers and districts, more training on what the test scores mean, and training in test score uses for policy makers. **More Support to Districts** includes greater assistance and more resources related to the learning standards and assistance in using test results for curriculum improvement. Several districts requested **Closer Alignment** between ISAT and the learning standards, specific links between questions and standards, and wider awareness of which standards are tested and which are not. Districts also requested **Improved Technical Quality,** including greater review of questions in the tests, better reliability, and greater involvement of both educators and technical experts. Several districts requested more open-ended and performance-based questions, and more opportunity for applied learning. On a related note, several districts suggested that the state turn the testing program over to a major commercial test publisher. Three districts stated that the ISAT was too difficult, that expectations were too high, and that the content needed to be "more realistic." Finally, two districts wanted a better alternative for students with disabilities.

Impact of an Improved ISAT

The final question in the survey asked districts to rate the impact that changes or improvements in the ISAT would have on their district testing programs. The scale ranged from 1 (not at all) to 5 (a great extent). As shown in Figure 8, most districts used the middle responses to describe the extent of changes they would make in response to improvement on the ISAT. There are slightly more responses on the positive end of the scale (that is, districts indicating they will make changes in their testing programs) than at the lower end (no or few changes), however the preponderance of responses in the middle suggests

Figure 8

Impact that Improvements to the ISAT Would Have on District Testing Program



widespread ambiguity about the effects that changes on the ISAT will have on district testing programs.

Summary Themes

A few core themes emerge across all of the different questions on this survey. The first is the importance and value that districts place on perceived quality and trustworthiness of tests. On the whole, they are very positive about the standardized tests that they purchase for their district testing programs, and much less favorable about the state testing program. Districts place a lot of faith in their own standardized tests and view them as highly trustworthy, reliable, and excellent sources of very useful information. They used words like "quality" and "integrity" in describing these tests. Districts were clearly less sanguine about the ISAT. Not only are they unhappy about the scheduling of the test and the turnaround time for scoring, but they comment on the need for greater consistency and stability in the state testing program. They would like to see the same quality in the ISAT that they see in their own standardized tests.

A second theme relates to the alignment between tests, and learning standards, and curriculum. Responses here are less straightforward. Though many districts would like better alignment between their own testing programs and learning standards, many are also content with measuring skills and knowledge that approximate rather than closely align to learning standards. Many districts also rate the ISAT positively for measuring state learning standards, though others suggest that the ISAT could be more closely aligned to the Illinois Learning Standards. This suggests the usefulness of assessments that provide an "external check" on measuring student performance as well as the more closely aligned assessment that provide information in relation to specific learning standards.

Finally, in several instances districts expressed willingness to make students "accountable" through test score results, while at the same time wishing to de-emphasize school level accountability. Are test results improvement tools or accountability tools? In the minds of school districts, there is much uncertainty on this issue.

This report reflects the interpretation of its authors. Although the Consortium assisted in the development of this research, no formal endorsement by its Steering Committee members, their organizations, or the Consortium should be assumed.